

ANALISIS SENTIMEN PADA REVIEW RESTORAN DENGAN TEKS BAHASA INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES

Dinda Ayu Muthia

Program Studi Manajemen Informatika

Akademi Manajemen Informatika dan Komputer Bina Sarana Informatika (AMIK BSI)

Jl. Cut Mutiah No. 88, Bekasi

dinda.dam@bsi.ac.id

ABSTRACT— *In the era of the web as it is now, some information is now flowing through the network. Because of the variety of web content includes subjective opinion and objective information, it is now common for people to gather information about products and services they want to buy. However, because there are quite a lot of information in text form without any numerical scale, it is difficult to classify the evaluation of information efficiently without reading the complete text. Sentiment analysis aims to address this problem by automatically classifying user review be positive or negative opinion. Naïve Bayes classifier is a popular machine learning techniques for text classification, because it is very simple, efficient and performs well in many domains. However, Naïve Bayes has the disadvantage that is very sensitive to feature too much, resulting in a classification accuracy becomes low. Therefore, in this study used the method of selecting features, namely Genetic algorithm in order to improve the accuracy of Naïve Bayes classifier. This research resulted in the classification of the text in the form of a positive or negative review of the restaurant. Measurement is based on the accuracy of Naive Bayes before and after the addition of feature selection methods. The evaluation was done using a 10 fold cross validation. While the measurement accuracy is measured by confusion matrix and ROC curves. The results showed an increase in the accuracy of Naïve Bayes from 86.50% to 90.50%.*

Keyword: *sentiment analysis, review, restaurant, naïve bayes, text classification.*

INTISARI— Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, sekarang umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap. Analisa sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan review

pengguna menjadi opini positif atau negatif. Pengklasifikasi *Naïve Bayes* adalah teknik *machine learning* yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, *Naïve Bayes* memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah. Oleh karena itu, dalam penelitian ini digunakan metode pemilihan fitur, yaitu *Genetic algorithm* agar bisa meningkatkan akurasi pengklasifikasi *Naïve Bayes*. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari review restoran. Pengukuran berdasarkan akurasi *Naive Bayes* sebelum dan sesudah penambahan metode pemilihan fitur. Evaluasi dilakukan menggunakan 10 *fold cross validation*. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi *Naïve Bayes* dari 86.50% menjadi 90.50%.

Kata Kunci: *analisis sentimen, review, restoran, naïve bayes, klasifikasi teks.*

PENDAHULUAN

1.1. Latar Belakang

Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, sekarang umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap (Lee, 2010).

Sudah ada beberapa penelitian analisis sentimen yang dilakukan untuk mengklasifikasi review ataupun opini restoran yang ada di internet di antaranya, klasifikasi sentimen pada review restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi *Naïve Bayes*

dan *Support Vector Machine* yang dilakukan oleh Z. Zhang pada tahun 2011. Analisis sentimen multiclass pada review restoran yang dilakukan oleh Moontae Lee pada tahun 2010. Kategorisasi sentimen pada review restoran yang dilakukan oleh Amir Ghazvinian pada tahun 2010.

Pada penelitiannya di tahun 2011, Zhang dan kawan-kawan meneliti penggunaan fitur n-gram dalam menggabungkan kata agar dapat dilihat perbedaan sentimen dari tiap gabungan kata. Rupanya penggunaan fitur n-gram berpengaruh dengan akurasi yang dihasilkan pengklasifikasi. Jika fitur yang digunakan 2-gram (gabungan 2 kata) atau 3-gram (gabungan 3 kata), maka otomatis jumlah kata yang diproses akan semakin banyak dan bisa mempengaruhi akurasi klasifikasi.

Pengklasifikasi *Naïve Bayes* sangat sederhana dan efisien (Chen, Huang, Tian, & Qu, 2009). Selain itu, pengklasifikasi *Naïve Bayes* adalah teknik *machine learning* yang populer untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain (Ye, Zhang, & Law, 2009). Namun, *Naïve Bayes* memiliki kekurangan yaitu sangat sensitif dalam pemilihan fitur (Chen et al., 2009). Terlalu banyak jumlah fitur, tidak hanya meningkatkan waktu penghitungan tapi juga menurunkan akurasi klasifikasi (Uysal & Gunal, 2012).

Hal lain yang ditemukan dalam analisis sentimen adalah pemilihan fitur. Pemilihan fitur bisa membuat pengklasifikasi baik lebih efektif dan efisien dengan mengurangi jumlah data yang dianalisa, maupun mengidentifikasi fitur yang sesuai untuk dipertimbangkan dalam proses pembelajaran. Ada dua jenis utama metode pemilihan fitur dalam *machine learning*: *wrapper* dan *filter*. *Wrapper* menggunakan akurasi klasifikasi dari beberapa algoritma sebagai fungsi evaluasinya. *Wrapper* mengevaluasi fitur secara berulang dan menghasilkan akurasi klasifikasi yang tinggi. Salah satu metode *wrapper* yang bisa digunakan dalam pemilihan fitur adalah *Genetic Algorithm* (GA).

Pada penelitian ini penggunaan fitur 2-gram dan 3-gram akan dibandingkan untuk mengetahui seberapa besar pengaruhnya terhadap akurasi klasifikasi dan pengklasifikasi *Naïve Bayes* dengan *Genetic Algorithm* sebagai metode pemilihan fitur akan diterapkan untuk mengklasifikasikan review restoran untuk meningkatkan akurasi analisis sentimen.

1.2. Identifikasi Masalah

Algoritma *Naïve Bayes* sangat sederhana, efisien dan merupakan teknik *machine learning* yang populer untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain. Namun, *Naïve Bayes* mempunyai kekurangan, yaitu sangat

sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah.

1.3. Rumusan Masalah

Seberapa besar efek metode pemilihan fitur *Genetic Algorithm* pada akurasi analisa sentimen pada review restoran dengan teks bahasa Indonesia menggunakan algoritma *Naïve Bayes*?

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengintegrasikan metode pemilihan fitur *Genetic Algorithm* dalam menganalisa sentimen pada review restoran dengan teks bahasa Indonesia menggunakan algoritma *Naïve Bayes*.

1.5. Manfaat Penelitian

1. Membantu para calon pembeli dalam mengambil keputusan saat ingin mencari restoran agar bisa mengurangi waktu dalam membaca review dan komentar dari suatu restoran.
2. Membantu para pengembang sistem yang berkaitan dengan review restoran, baik dari sumber Zomato maupun dari sosial media lainnya seperti Twitter, Blog, dan lain-lain.

1.6. Kontribusi Penelitian

Mengklasifikasikan teks bahasa Indonesia untuk menganalisa sentimen pada review restoran dengan menggunakan algoritma *Naïve Bayes* yang menerapkan metode pemilihan fitur *Genetic Algorithm*.

BAHAN DAN METODE

2.1. Review

Menurut Reddy (V, Somayajulu, & Dani, 2010) di beberapa kasus, keputusan yang kita buat dipengaruhi oleh opini atau pendapat dari orang lain. Sebelum kesadaran akan internet menjadi tersebar luas, banyak dari kita yang biasanya menanyakan opini teman atau tetangga mengenai alat elektronik ataupun film sebelum benar-benar membelinya. Dengan berkembangnya ketersediaan dan popularitas akan sumber yang kaya opini seperti website review online dan blog pribadi, kesempatan baru dan tantangan muncul semenjak orang-orang sekarang bisa menggunakan informasi dan teknologi secara aktif untuk mencari dan memahami opini orang lain.

2.2. Analisis Sentimen

Menurut Feldman (Feldman, 2013) analisis sentimen adalah tugas menemukan opini dari penulis tentang suatu entitas tertentu. Menurut Tang dalam Haddi (Haddi, Liu, & Shi, 2013), analisis sentimen pada review adalah proses menyelidiki

review produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan.

Menurut Thelwall dalam Haddi, analisa sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif.

2.3. Algoritma Naive Bayes

Menurut Markov (Markov & Daniel, 2007) tahapan dalam algoritma Naive Bayes:

1. Hitung probabilitas bersyarat/likelihood:

$$P(x | C) = P(x_1, x_2, \dots, x_n | C) \dots\dots\dots (1)$$

$$C = \text{class}$$

$$x = \text{vektor dari nilai atribut } n$$

$$P(x_i | C) = \text{proporsi dokumen dari class } C \text{ yang mengandung nilai atribut } x_i$$
2. Hitung probabilitas prior untuk tiap class:

$$P(C) = \frac{N_j}{N} \dots\dots\dots (2)$$

$$N_j = \text{jumlah dokumen pada suatu class}$$

$$N = \text{jumlah total dokumen}$$
3. Hitung probabilitas posterior dengan rumus:

$$P(C|x) = \frac{P(x|C) P(C)}{P(x)} \dots\dots\dots (3)$$

Menurut Santoso (Santoso, 2007) dengan kata-kata yang lebih umum, rumus Bayes bisa diberikan sebagai berikut:

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} \dots\dots\dots (4)$$

2.4. Pemilihan Fitur

Menurut Gorunescu (Gorunescu, 2011) pemilihan fitur digunakan untuk menghilangkan fitur yang tidak relevan dan berulang, yang mungkin menyebabkan kekacauan, dengan menggunakan metode tertentu. Menurut John, Kohavi, dan Pfleger dalam Chen (Chen et al., 2009) ada dua jenis metode pemilihan fitur dalam machine learning, yaitu wrapper dan filter.

Menurut Chen (Chen et al., 2009) wrapper menggunakan akurasi klasifikasi dari beberapa algoritma sebagai fungsi evaluasinya. Menurut Gunal (Gunal, 2012) salah satu metode wrapper yang bisa digunakan dalam pemilihan fitur adalah *Genetic Algorithm* (GA).

Menurut Han (Han & Kamber, 2007) *Genetic Algorithm* berusaha untuk menggabungkan ide-ide evolusi alam. Secara umum, pembelajaran genetika dimulai sebagai berikut:

- a. Sebuah populasi awal dibuat terdiri dari aturan acak. Setiap aturan bisa diwakili oleh string bit. Sebagai contoh sederhana, misalkan bahwa sampel dalam satu set pelatihan yang diberikan dijelaskan oleh dua atribut Boolean, A1 dan A2, dan bahwa ada dua kelas, C1 and C2. Aturan "If A1 And Not A2 Then C2" dapat dikodekan sebagai string bit "100," di mana dua bit paling kiri

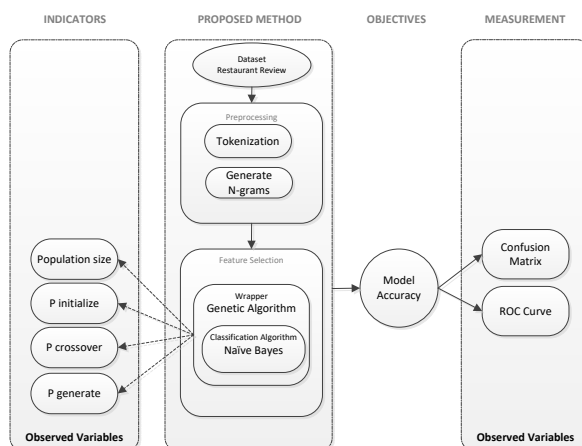
mewakili atribut A1 dan A2, masing-masing, dan bit paling kanan mewakili kelas. Demikian pula, aturan "If Not A1 And Not A2 Then C1" dapat dikodekan sebagai "001." Jika atribut memiliki nilai-nilai k, di mana $k > 2$, maka k bit dapat digunakan untuk mengkodekan nilai-nilai atribut itu. Kelas dapat dikodekan dengan cara yang sama.

- b. Berdasarkan gagasan ketahanan dari yang paling sesuai, populasi baru terbentuk terdiri dari aturan yang paling sesuai dalam populasi saat ini, serta keturunan aturan ini. Biasanya, fitness aturan dinilai dengan akurasi klasifikasi pada satu set sampel pelatihan.
- c. Keturunan diciptakan dengan menerapkan operator genetika seperti crossover dan mutasi. Dalam crossover, substring dari sepasang aturan ditukar untuk membentuk pasangan aturan baru. Dalam mutasi, bit yang dipilih secara acak dalam aturan string dibalik.
- d. Proses menghasilkan populasi baru berdasarkan aturan populasi sebelumnya berlanjut sampai populasi, P, berkembang di mana setiap aturan dalam P memenuhi ambang batas fitness yang sudah ditentukan.

Genetic Algorithm mudah disejajarkan dan telah digunakan untuk klasifikasi seperti masalah optimasi lainnya. Dalam data mining, algoritma genetika dapat digunakan untuk mengevaluasi fitness algoritma lainnya.

2.5. Kerangka Pemikiran

Penelitian ini dimulai dari adanya masalah dalam klasifikasi teks pada review film menggunakan *Naive Bayes*, di mana pengklasifikasi tersebut memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah. Dataset yang digunakan dalam penelitian ini berupa review restoran dengan teks bahasa Indonesia yang didapat dari situs khusus review restoran yaitu www.zomato.com yang terdiri dari 100 review positif dan 100 review negatif. *Preprocessing* yang dilakukan dengan *tokenization* dan *generate N-grams*. Metode pemilihan fitur yang digunakan adalah *Genetic Algorithm*, sedangkan pengklasifikasi yang digunakan adalah *Naive Bayes*. Pengujian 10 *fold cross validation* akan dilakukan, akurasi algoritma akan diukur menggunakan *confusion matrix* dan hasil olahan data dalam bentuk kurva ROC. RapidMiner versi 7.3 digunakan sebagai alat bantu dalam mengukur akurasi data eksperimen. Gambar 1. menggambarkan kerangka pemikiran yang penulis usulkan dalam penelitian ini.



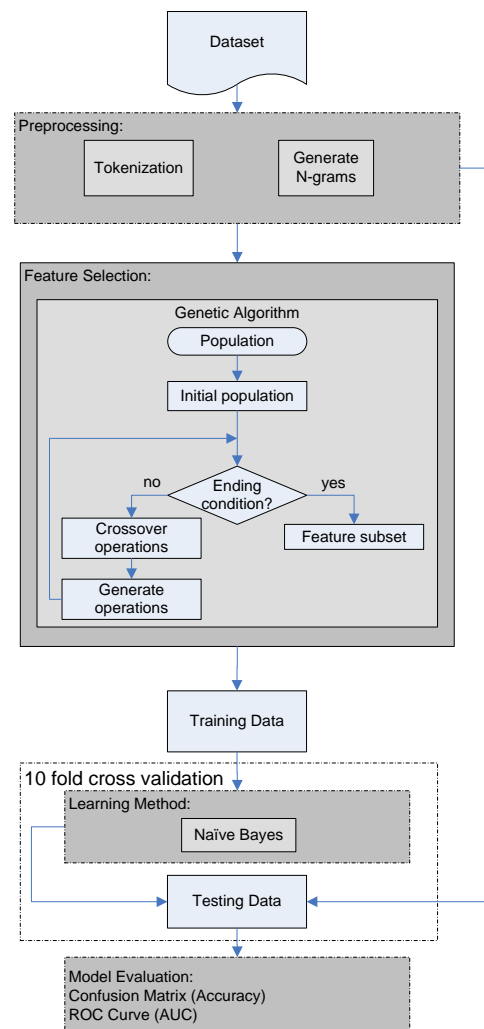
Sumber: Hasil Penelitian (2016)
Gambar 1. Kerangka Pemikiran

2.6. Metode Penelitian

Metode penelitian yang penulis lakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut:

- a. **Pengumpulan Data**
Penulis menggunakan data review restoran yang didapat dari situs www.zomato.com yang terdiri dari 100 review positif dan 100 review negatif.
- b. **Pengolahan Awal Data**
Dataset ini dalam tahap *preprocessing* harus melalui 2 proses, yaitu:
 - 1) **Tokenization**
Yaitu mengumpulkan semua kata yang muncul dan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.
 - 2) **Generate N-grams**
Yaitu menggabungkan kata sifat yang seringkali muncul untuk menunjukkan sentimen, seperti kata "sangat" dan kata "bagus". Kata "bagus" memang sudah menunjukkan sentimen bentuk opini positif. Kata "sangat" tidak akan berarti jika berdiri sendiri. Namun jika dua kata tersebut digabung menjadi "sangat bagus", maka akan sangat menguatkan opini positif tersebut. Penulis menggunakan penggabungan tiga kata, yang disebut 3-grams (Trigrams).
- c. **Metode yang Diusulkan**
Metode yang penulis usulkan adalah menggabungkan *Genetic Algorithm* sebagai metode pemilihan fitur agar akurasi pengklasifikasi *Naïve Bayes* bisa meningkat. Penulis menggunakan pengklasifikasi *Naïve Bayes* karena sangat sederhana, efisien dan merupakan teknik *machine learning* yang populer untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain. *Genetic Algorithm* yang penulis terapkan adalah menggunakan *Naïve Bayes* yang diuji di

dalam tahap wrapper. Lihat gambar 2. untuk model yang diusulkan secara lebih detail.



Sumber: Hasil Penelitian (2016)
Gambar 2. Model yang diusulkan

HASIL DAN PEMBAHASAN

3.1. Hasil

3.1.1. Klasifikasi Teks Menggunakan Algoritma Naïve Bayes

Data training yang digunakan dalam pengklasifikasian teks ini terdiri dari 100 review restoran positif dan 100 review restoran negatif. Data tersebut masih berupa sekumpulan teks yang terpisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melalui beberapa tahapan proses agar bisa diklasifikasikan dalam proses selanjutnya, berikut adalah tahapan prosesnya:

1. **Pengumpulan Data**
Data review positif disatukan dalam folder dengan nama pos. Sedangkan data review negatif disatukan penyimpanannya dalam

folder dengan nama neg. Tiap dokumen berekstensi .txt yang dapat dibuka menggunakan aplikasi Notepad.

2. Pengolahan Awal Data
Proses yang dilalui terdiri dari *tokenization*, dan *generate 3-grams*.
3. Klasifikasi
Proses klasifikasi di sini adalah untuk menentukan sebuah kalimat sebagai anggota class positif atau class negatif berdasarkan nilai perhitungan probabilitas dari rumus Bayes yang lebih besar. Jika hasil probabilitas kalimat tersebut untuk class positif lebih besar dari pada class negatif, maka kalimat tersebut termasuk ke dalam class positif. Jika probabilitas untuk class positif lebih kecil dari pada class negatif, maka kalimat tersebut termasuk ke dalam class negatif. Penulis mendapatkan 3 kata dengan sentimen positif yang paling sering muncul, yaitu "enak", "ramah", dan "nyaman". Sedangkan kata dengan sentimen negatif yang sering muncul adalah "mahal", "kecewa", dan ditemukan hasil dari *generate N-Grams* yang sering muncul adalah gabungan kata "agak kurang".

3.1.2. Pengujian Model dengan 10 Fold Cross Validation

Pada penelitian ini, penulis melakukan pengujian model dengan menggunakan teknik 10 *cross validation*, di mana proses ini membagi data secara acak ke dalam 10 bagian. Proses pengujian dimulai dengan pembentukan model dengan data pada bagian pertama. Model yang terbentuk akan diujikan pada 9 bagian data sisanya. Setelah itu proses akurasi dihitung dengan melihat seberapa banyak data yang sudah terklasifikasi dengan benar.

3.1.3. Eksperimen Terhadap Indikator Model

Untuk mendapatkan model yang baik, beberapa indikator disesuaikan nilainya agar didapatkan hasil akurasi yang tinggi. Dalam penyesuaian indikator pada Genetic algorithm, akurasi paling tinggi diperoleh dengan kombinasi $p_{initialize}=0.8$ dan $p_{crossover}=0.8$. Hasil akurasi mencapai 90.50% dari yang sebelumnya akurasi bernilai 86.50%. Jika indikator lainnya turut diubah nilainya, dapat menyebabkan proses pengolahan data menjadi semakin lama.

3.2. Pembahasan

Dengan memiliki model klasifikasi teks pada review, pembaca dapat dengan mudah mengidentifikasi mana review yang positif maupun yang negatif. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen

yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut positif atau negatif.

Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan seberapa baik model yang terbentuk. Tanpa menggunakan metode pemilihan fitur, algoritma *Naïve Bayes* sendiri sudah menghasilkan akurasi sebesar 86.50%. Akurasi tersebut masih perlu ditingkatkan lagi menggunakan metode pemilihan fitur. Setelah menggunakan metode pemilihan fitur dari wrapper yang digabungkan, akurasi algoritma *Naïve Bayes* meningkat menjadi 90.50% seperti yang bisa dilihat pada tabel 3.

Tabel 1. Model algoritma *Naïve Bayes* sebelum dan sesudah menggunakan metode pemilihan fitur

| | Algoritma <i>Naïve Bayes</i> | Algoritma <i>Naïve Bayes + Genetic Algorithm</i> |
|-----------------------------------|------------------------------|--|
| Sukses klasifikasi review positif | 94 | 96 |
| Sukses klasifikasi review negatif | 79 | 85 |
| Akurasi model | 86.50% | 90.50% |

Sumber: Hasil Penelitian (2016)

3.2.1. Pengukuran dengan *Confusion Matrix*

Pengukuran dengan *confusion matrix* di sini akan menampilkan perbandingan dari hasil akurasi model *Naïve Bayes* sebelum ditambahkan metode pemilihan fitur yang bisa dilihat pada tabel 2 dan setelah ditambahkan metode pemilihan fitur, yaitu *Genetic algorithm* yang bisa dilihat pada tabel 3.

Tabel 2. *Confusion matrix* model *Naïve Bayes* sebelum penambahan metode pemilihan fitur

| Akurasi <i>Naïve Bayes</i> : 86.50% +/- 7.43% (mikro: 86.50%) | | | |
|---|--------------|--------------|-----------------|
| | True Positif | True Negatif | Class precision |
| Pred. Positif | 94 | 21 | 81.74% |
| Pred. Negatif | 5 | 79 | 92.94% |
| Class recall | 94.00% | 79.00% | |

Sumber: Hasil Penelitian (2016)

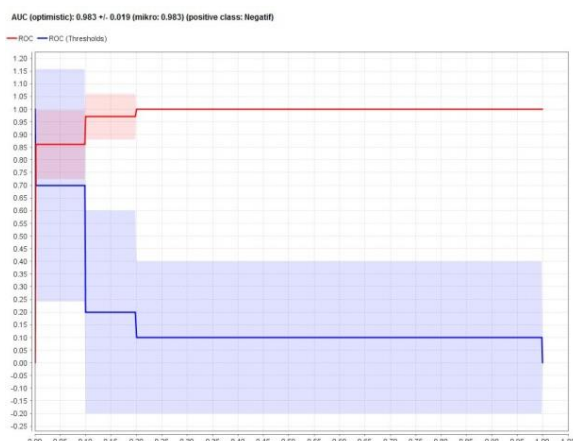
Tabel 3. *Confusion matrix* model *Naïve Bayes* sesudah penambahan metode pemilihan fitur

Akurasi *Naive Bayes*: 84.50% +/- 5.22%
(mikro: 84.50%)

| | True Positif | True Negatif | Class precision |
|---------------|--------------|--------------|-----------------|
| Pred. Positif | 96 | 15 | 86.49% |
| Pred. Negatif | 4 | 85 | 95.51% |
| Class recall | 96.00% | 85.00% | |

Sumber: Hasil Penelitian (2016)

Berikut adalah tampilan kurva ROC dari hasil uji data. Gambar 3 adalah kurva ROC untuk model *Naive Bayes* sebelum menggunakan metode pemilihan fitur dan gambar 4 adalah kurva ROC untuk model *Naive Bayes* setelah menggunakan metode pemilihan fitur.



Sumber: Hasil Penelitian (2016)

Gambar 3. Kurva ROC model *Naive Bayes* sebelum menggunakan metode pemilihan fitur



Sumber: Hasil Penelitian (2016)

Gambar 4. Kurva ROC model *Naive Bayes* setelah menggunakan metode pemilihan fitur

3.3. Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

1. Implikasi terhadap aspek sistem
Hasil evaluasi menunjukkan penerapan *Genetic Algorithm* untuk seleksi fitur dapat meningkatkan akurasi *Naive Bayes* dan merupakan metode yang cukup baik dalam mengklasifikasi review restoran dengan teks bahasa Indonesia. Dengan demikian penerapan metode tersebut dapat membantu para calon pembeli dalam mengambil keputusan saat ingin mencari restoran yang sesuai dengan keinginannya.
2. Implikasi terhadap aspek manajerial
Membantu para pengembang sistem yang berkaitan dengan review restoran, baik dari sumber www.zomato.com maupun dari sosial media lainnya seperti Twitter, Blog, dan lain-lain agar menggunakan aplikasi RapidMiner dalam membangun suatu sistem.
3. Implikasi terhadap aspek penelitian lanjutan
Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti review produk, review travel, dan sebagainya.

KESIMPULAN

Dari pengolahan data yang sudah dilakukan, *Genetic Algorithm* terbukti dapat meningkatkan akurasi pengklasifikasi *Naive Bayes*. Data review restoran dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi *Naive Bayes* sebelum menggunakan penggabungan metode pemilihan fitur mencapai 86.50%. Sedangkan setelah menggunakan penggabungan metode pemilihan fitur, yaitu *Genetic Algorithm*, akurasinya meningkat hingga mencapai 90.50%. Peningkatan akurasi mencapai 4%.

Model yang terbentuk dapat diterapkan pada seluruh data review restoran, sehingga dapat dilihat secara langsung hasilnya dalam bentuk positif dan negatif. Hal ini dapat membantu seseorang untuk menghemat waktu saat mencari review suatu restoran tanpa harus mengkhawatirkan pemberian rating yang tidak sesuai dengan reviewnya.

REFERENSI

- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with *Naive Bayes*. *Expert Systems with Applications*, 36(3), 5432–5435.
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82.

- Gorunescu, F. (2011). *Data Mining Concept Model Technique*.
- Gunal, S. (2012). Hybrid feature selection for text classification ", 20.
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32.
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*.
- Lee, M. (2010). MULTICLASS SENTIMENT ANALYSIS WITH RESTAURANT REVIEWS.
- Markov, Z., & Daniel, T. (2007). *Uncovering Patterns in*.
- Santoso, Budi. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Uysal, A. K., & Gunal, S. (2012). A novel probabilistic feature selection method for text classification. *Knowledge-Based Systems*, 36, 226–235.
- V, S. R. R., Somayajulu, D. V. L. N., & Dani, A. R. (2010). Classification of Movie Reviews Using Complemented Naive Bayesian Classifier, 1(4), 162–167.
- Ye, Q., Zhang, Z., & Law, R. (2009). Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems With Applications*, 36(3), 6527–6535.

BIODATA PENULIS



Dinda Ayu Muthia, M.Kom. Lahir di Jakarta, 8 Juli 1988. Tahun 2011 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta. Tahun 2013 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Tahun 2015 sudah memiliki Jabatan Fungsional Akademik Asisten Ahli di AMIK BSI Jakarta. Jurnal yang pernah diterbitkan di antaranya, Analisis Sentimen Pada Review Buku Menggunakan Algoritma *Naïve Bayes* pada tahun 2014 di Jurnal Paradigma dan Opinion Mining Pada Review Buku Menggunakan Algoritma *Naïve Bayes* di Jurnal Teknik Komputer pada tahun 2016. Proceeding yang pernah diseminarkan di ISSIT 2014 dengan judul *Sentiment Analysis of Hotel Review Using Naive Bayes Algorithm and Integration of Information Gain and Genetic Algorithm as Feature Selections Methods*.